

(19) RÉPUBLIQUE FRANÇAISE
INSTITUT NATIONAL
DE LA PROPRIÉTÉ INDUSTRIELLE
PARIS

(11) N° de publication :
(à n'utiliser que pour les
commandes de reproduction)

2 659 461

(21) N° d'enregistrement national :

90 03071

(51) Int Cl⁵ : G 06 F 15/38

(12)

DEMANDE DE BREVET D'INVENTION

A1

(22) Date de dépôt : 06.03.90.

(30) Priorité :

(71) Demandeur(s) : KIS FRANCE (société à
responsabilité limitée) — FR et CRASNIANSKI Serge
— FR.

(72) Inventeur(s) : CRASNIANSKI Serge et Gruel Henri.

(43) Date de la mise à disposition du public de la
demande : 13.09.91 Bulletin 91/37.

(56) Liste des documents cités dans le rapport de
recherche : Se reporter à la fin du présent fascicule.

(60) Références à d'autres documents nationaux
apparentés :

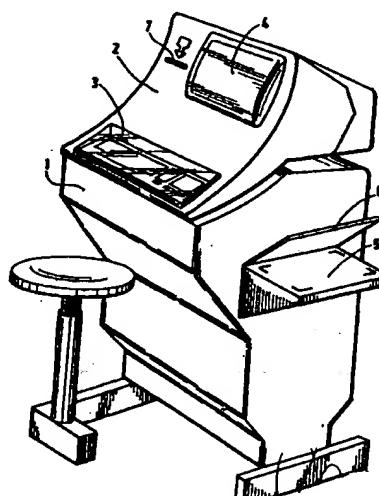
(73) Titulaire(s) :

(74) Mandataire : Cabinet Laurent.

(54) Dispositif de traduction automatique d'un document.

(57) Dispositif pour effectuer automatiquement une traduction
d'un document, caractérisé en ce qu'il comprend:

- un scanner (5);
- une unité centrale de traitement (10) constituée par un ordinateur, reliée au scanner (5), et dont la mémoire stocke:
 - . d'une part, un logiciel de reconnaissance de caractères;
 - . d'autre part, un logiciel de traduction dans la langue considérée à partir des caractères reconnus et de leur séquencement;
- un écran (4) relié à l'unité centrale (10) permettant la visualisation du processus de traduction et les propositions de traduction émises par l'ordinateur (10);
- un clavier (3), relié à l'unité centrale (10), pour initialiser le processus de traduction et valider les propositions de traduction émises par l'ordinateur et affichées à l'écran (4);
- une imprimante (11) reliée à l'unité centrale (10) pour restituer la traduction proposée.



FR 2 659 461 - A1



DISPOSITIF DE TRADUCTION AUTOMATIQUE D'UN DOCUMENT.

L'invention a pour but de proposer un dispositif apte à permettre la réalisation d'une traduction d'un document, notamment d'un document dactylographié.

A ce jour, on ne connaît pas de dispositif permettant d'obtenir une traduction intégrale d'un texte même dactylographié. Il existe certes de petits micro-ordinateurs, intégrant un logiciel de traduction et permettant de restituer de manière immédiate la traduction ponctuelle d'un mot entré par l'utilisateur. Or, dans le cadre de l'échange international sans cesse croissant, il s'avère nécessaire de réaliser une machine susceptible de pouvoir restituer une traduction intégrale d'un document dans un délai raisonnable.

C'est l'objet de la présente invention. Celle-ci propose un dispositif pour effectuer automatiquement la traduction d'un document, caractérisé en ce qu'il comprend :

- un scanner comportant :
 - . une plaque transparente sur laquelle on applique la feuille du document à traduire ;
 - . et un chariot mobile en translation présentant une pluralité de photorécepteurs destinés à capter les signaux lumineux réfléchis par le document éclairé au moyen d'une source lumineuse, et à transformer ces signaux lumineux en signaux analogiques, puis en signaux numériques ;
 - une unité centrale de traitement constituée par un ordinateur, reliée au scanner, et dont la mémoire stocke :

- d'une part, un logiciel de reconnaissance de caractères à partir des signaux numériques émis par le scanner ;
- et d'autre part, un logiciel de traduction dans la langue considérée à partir des caractères reconnus et de leur séquencement ;

5 - un écran relié à l'unité centrale, permettant la visualisation du processus de traduction et les propositions de traduction réalisées par l'ordinateur ;

10 - un clavier, relié à l'unité centrale, afin d'initialiser le processus de traduction et de valider les propositions de traduction émises par l'ordinateur et affichées sur l'écran ;

15 - une imprimante, reliée à l'unité centrale, afin de restituer la traduction proposée.

En d'autres termes, l'invention consiste essentiellement à associer dans le cadre d'une architecture particulière, un logiciel de reconnaissance de caractères et 20 un logiciel de traduction, en vue de permettre à partir d'un texte donné, la restitution intégrale du texte dans la langue choisie.

Avantageusement, en pratique :

25 - le déplacement en translation du chariot du scanner est commandé par un moteur pas à pas ; le pas choisi peut soit être fonction de l'interligne reconnu du texte dactylographié à traduire, soit un pas préalablement établi lors de la mise en place du logiciel de reconnaissance de caractères, ce dernier étant alors capable de restituer la continuité du document d'un balayage linéaire au suivant ;

30 - l'ordinateur est apte à reconnaître les formes de présentation du texte à traduire, et à éliminer les caractères non graphiques ;

35

- le dispositif comprend en outre un organe de pré-paiement relié au moyen d'un interface à l'unité centrale ;

- cet organe de pré-paiement est un monnayeur ou un 5 lecteur de cartes, voire de billets de banque.

La manière dont l'invention peut être réalisée et les avantages qui en découlent ressortiront mieux de l'exemple de réalisation qui suit, donné à titre indicatif et non limitatif, à l'appui des figures annexées. 10

La figure 1 est un schéma synoptique simplifié de la structure du dispositif conforme à l'invention.

La figure 2 est une vue schématique en perspective 15 d'une forme de réalisation du dispositif conforme à l'invention.

Selon l'invention, le dispositif apte à réaliser une traduction intégrale d'un document, comprend essentiellement un bâti (1) recevant une console (2), sur 20 lequel sont agencés les différents éléments nécessaires au contrôle et à l'obtention de la traduction. Associé à cette console, on ménage un scanner (5) comportant une vitre transparente horizontale sur laquelle on applique 25 la face du document à traduire. De manière connue, on rabat un couvercle (6) sur cette vitre (5), système que l'on rencontre couramment dans les photocopies. Ce scanner (5) comporte une source lumineuse située sous la vitre (5), et destinée à éclairer la face du document à 30 traduire. Il comprend en outre une barrette de photodiodes destinée à se déplacer en translation afin de balayer toute la face dudit document. Les photodiodes sont destinées à capter les signaux lumineux réfléchis par la face éclairée et à transformer ainsi ces signaux lumineux en 35 signaux analogiques. Dans une forme de réalisation avan-

tageuse, ces signaux analogiques sont transformés en signaux numériques ou en niveaux de gris (pixels). De la sorte, on obtient une cartographie numérisée correspondant et représentative de la face imprimée du document à traduire. La résolution typique de ce scanner est voisine de douze points par millimètre.

Cette cartographie numérisée est alors transmise et stockée dans la mémoire d'une unité centrale comportant fondamentalement un ordinateur (10). Ce dernier stocke en mémoire tout d'abord un logiciel apte à reconnaître les caractères graphiques à partir de la cartographie numérisée transmise par le scanner (5). Pour ce faire, ce logiciel procède tout d'abord à l'analyse de la page, et notamment à la reconnaissance du format. Il sélectionne alors parmi l'ensemble des différentes zones du document, les zones de texte qui seules seront analysées. Les éléments continus de ces zones sont alors détectés en vue de pouvoir être reconnu comme caractère lors de la procédure d'analyse.

Cette analyse de caractères commence tout d'abord par le calcul des alignements, et des paramètres propres à chaque document, tels que l'espacement et l'interligne. La reconnaissance proprement dite des caractères s'effectue par le biais d'un modèle mathématique, utilisant des concepts de reconnaissance topologique de formes .

A la fin de l'opération, les différents mots non reconnus sont affichés sur un écran (4), décrit ultérieurement, afin que l'utilisateur précise les caractères mal ou non-reconnus.

On obtient ainsi un fichier-résultat renfermant l'ensemble des mots reconnus du document dactylographié

de base, fichier qui va servir de base au logiciel de traduction, également stocké dans la mémoire de l'unité centrale (10). Fondamentalement, ce logiciel est destiné à traduire les mots, séquences de mots ou phrases reconnus par le logiciel de reconnaissance de caractères.

Le logiciel de traduction effectue préalablement une analyse grammaticale des phrases mises en mémoire dans le fichier-résultat. Il reconnaît en outre la forme des 10 phrases, à savoir affirmative, négative, interrogative, dont il connaît la forme correspondante dans la langue dans laquelle le document de base doit être traduit.

Le logiciel intègre notamment deux dictionnaires, à 15 savoir :

- un dictionnaire d'expressions courantes, constituées d'une suite de mots formant un tout "logique", intraduisibles mot à mot ;
- un dictionnaire de mots unitaires, de capacité 20 variable, mais au moins égale à 50 000 mots avec toutes leurs formes dérivées.

Ces deux dictionnaires sont susceptibles d'être enrichis par l'utilisateur. De manière avantageuse, ce 25 dernier peut opter préalablement au travail de ce logiciel, pour un domaine particulier, sélectionnant alors le dictionnaire correspondant.

Pour le cas des mots inconnus par le logiciel, un 30 échange interactif s'établit entre celui-ci et l'utilisateur, qui, dans la mesure de ses connaissances, doit fournir la traduction des mots manquants, et ainsi enrichir les dictionnaires.

Cette interactivité est rendue possible par l'utilisation d'un clavier (3) et d'un écran (4), disposés sur la console (2), et reliés à l'ordinateur. Le clavier est destiné tout d'abord à initialiser le processus de traduction et à valider les propositions de traduction présentées par l'ordinateur (10) et matérialisées sur l'écran de visualisation (4). De fait, pour chaque phrase ou séquence de mots, l'ordinateur (10) propose une ou plusieurs traductions, dont l'une est choisie puis validée par l'utilisateur. En outre, lorsque l'on désire procéder à la traduction d'un texte manuscrit, le clavier peut également servir à introduire ce texte en mémoire dans l'ordinateur (10).

15 Dans une réalisation du dispositif conforme à l'invention destinée à être proposée au public, on peut munir ce dernier d'un organe de pré-paiement constitué soit d'un monnayeur (7), soit d'un lecteur de cartes bancaires ou de pré-paiement, voire même d'un lecteur de billets 20 bancaires. Cet organe de pré-paiement (7) est relié à l'ordinateur (10) par l'intermédiaire d'un interface (12) apte à permettre, de manière connue, la détermination du montant acquitté.

25 Enfin, afin de restituer la traduction, le dispositif est équipé d'une imprimante (11) reliée à l'ordinateur (10), typiquement une imprimante à laser. Mais il va de soi que tout autre type d'imprimante est utilisable.

30

Lorsque l'on désire procéder à la traduction d'un document, on appose la face imprimée du document sur la vitre du scanner (5), on rabat le couvercle (6) sur la vitre, puis on procède à l'initialisation du processus de traduction. Cette initialisation, commandée par le

clavier (3), commence par le balayage du texte à traduire par la barette de photodiodes, puis se poursuit par l'obtention d'une cartographie numérisée transmise à l'ordinateur (10) et stockée dans celui-ci. Ce dernier 5 procède alors aux propositions de traduction affichées sur l'écran (4), puis validées par l'utilisateur. L'impression de la traduction peut s'effectuer ligne par ligne ou phrase par phrase, voire même en une seule étape, les différents choix validés de l'utilisateur 10 ayant été alors stockés dans une autre mémoire de l'ordinateur (10).

Ce dispositif permet ainsi de réaliser un gain de temps appréciable, en fournissant dans un délai raisonnable 15 une traduction d'un texte dactylographié dans la langue désirée. Pour cela, il suffit de préselectionner au départ le logiciel de traduction correspondant, et avantageusement le domaine technique, juridique, etc. le plus approprié.

REVENDICATIONS

1/ Dispositif pour effectuer automatiquement une traduction d'un document, caractérisé en ce qu'il com-
5 prend :

- un scanner (5) comportant :

- une plaque transparente sur laquelle on appli-
que la face du document à traduire ;
- et un chariot mobile en translation présentant
10 une pluralité de photorécepteurs destinés à capter les signaux lumineux réfléchis par le document éclairé par une source lumineuse interne, et à transformer ces signaux lumineux en signaux analogiques, puis en signaux numéri-
15 ques ;

- une unité centrale de traitement (10) constituée par un ordinateur, reliée au scanner (5), et dont la mémoire stocke :

- d'une part, un logiciel de reconnaissance de caractères à partir des signaux numériques émis par le scanner (5) ;
- d'autre part, un logiciel de traduction dans la langue considérée à partir des caractères re-
connus et de leur séquencement ;

25 - un écran (4) relié à l'unité centrale (10) permettant la visualisation du processus de traduction et les propositions de traduction émises par l'ordinateur (10) ;

- un clavier (3), relié à l'unité centrale (10), pour initialiser le processus de traduction et valider
30 les propositions de traduction émises par l'ordinateur et affichées à l'écran (4) ;

- une imprimante (11) reliée à l'unité centrale (10) pour restituer la traduction proposée.

2/ Dispositif selon la revendication 1, caractérisé en ce que le déplacement en translation du chariot du scanner (5) est commandé par un moteur pas à pas.

5 3/ Dispositif selon l'une des revendications 1 et 2, caractérisé en ce que l'ordinateur est apte à reconnaître les formes de présentation du texte à traduire et à éliminer les caractères non graphiques.

10 4/ Dispositif selon l'une des revendications 1 à 3, caractérisé en ce qu'il comprend également un organe de pré-paiement (7) relié à l'unité centrale (10) au moyen d'un interface (12), apte à convertir les signaux émis par ledit organe (7) en signaux compatibles pour l'ordinateur (10).

5/ Dispositif selon la revendication 4, caractérisé en ce que l'organe de pré-paiement (7) est un monnayeur.

20 6/ Dispositif selon la revendication 4, caractérisé en ce que l'organe de pré-paiement (7) est un lecteur de cartes ou de billets bancaires.

1/2

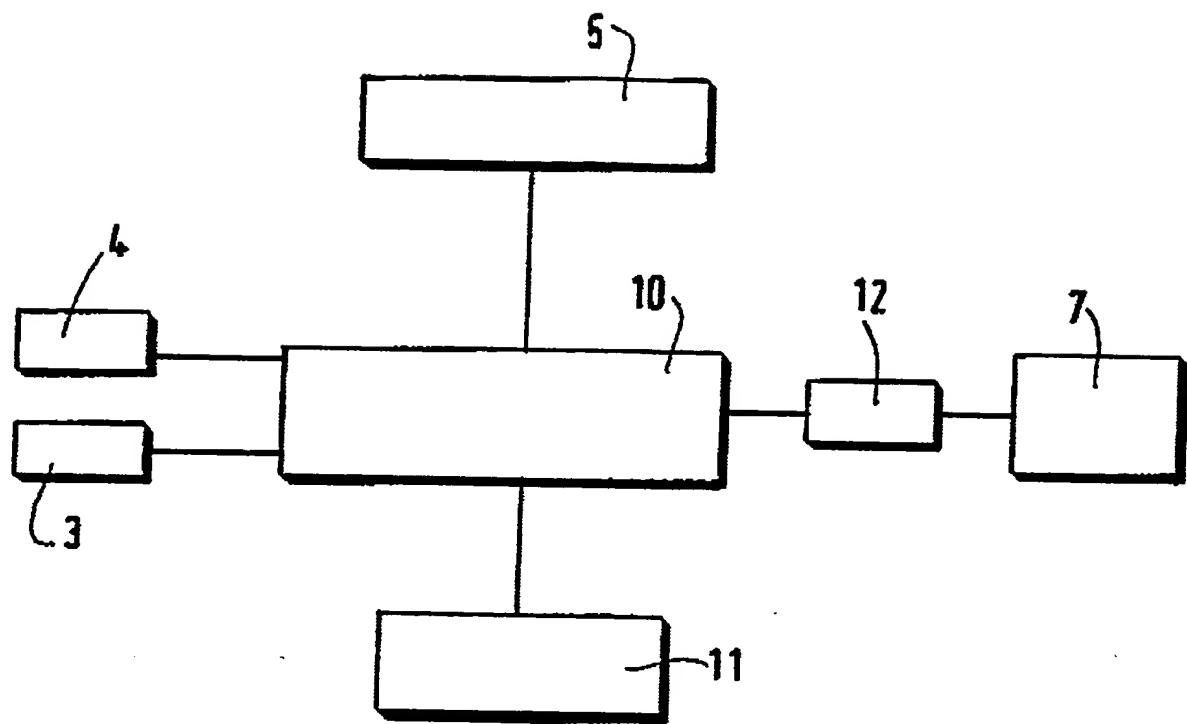


FIG.1

2/2

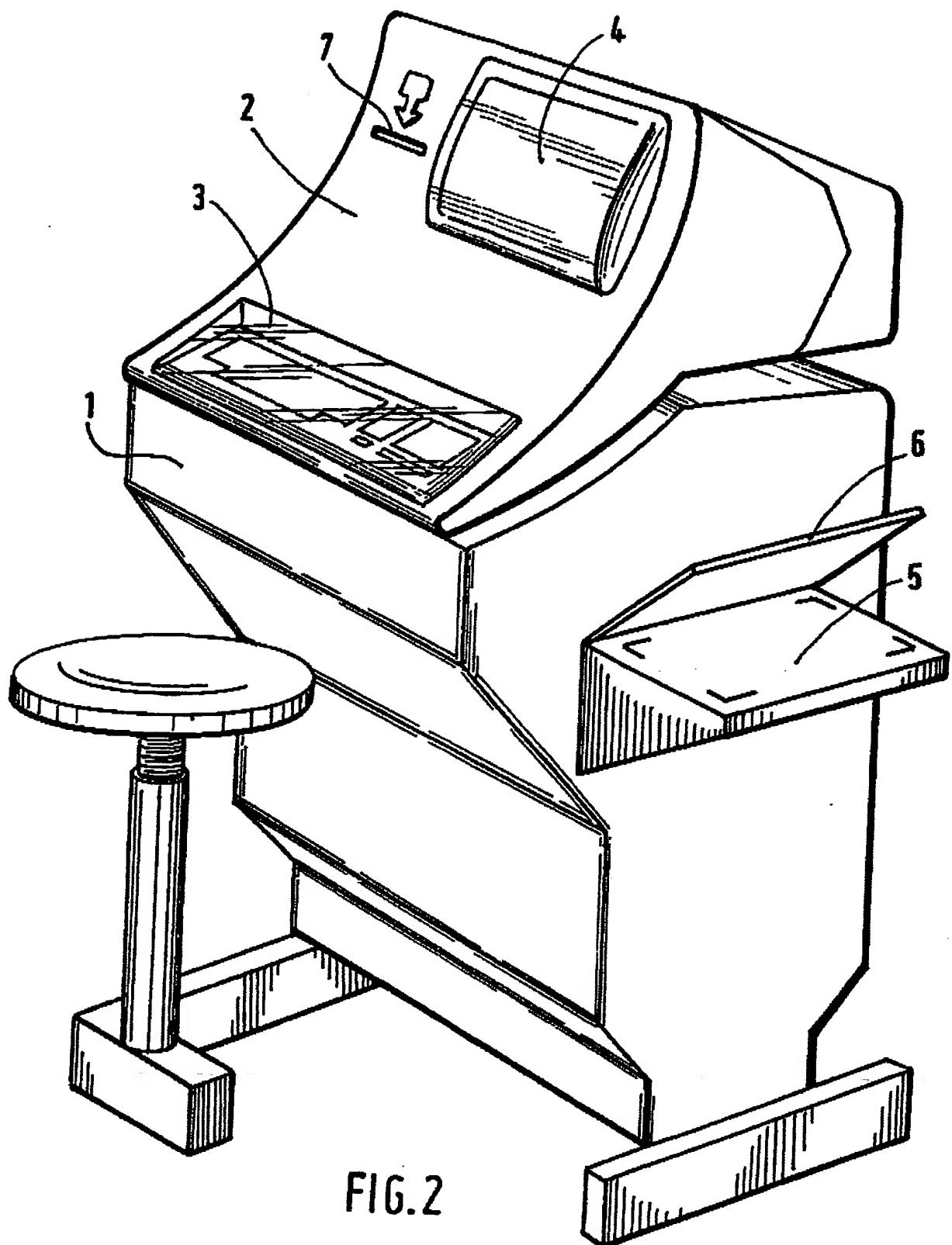


FIG.2

INSTITUT NATIONAL
de la
PROPRIETE INDUSTRIELLE

RAPPORT DE RECHERCHE

établi sur la base des dernières revendications
déposées avant le commencement de la rechercheFR 9003071
FA 444898

DOCUMENTS CONSIDERES COMME PERTINENTS		Revendications concernées de la demande examinée
Catégorie	Citation du document avec indication, en cas de besoin, des parties pertinentes	
X	PATENT ABSTRACTS OF JAPAN, vol. 9, no. 76 (P-346)[1799], 5 avril 1985; & JP-A-59 206 985 (HITACHI) * Abrégé * ---	1,3
X	EP-A-0 272 158 (OKI ELECTRIC) * Abrégé; colonne 1, lignes 37-55; colonne 2, lignes 1-10,14-16,21-24; colonne 4, lignes 34-39,51-53; colonne 5, lignes 2-5; colonne 6, lignes 1-9; colonne 7, lignes 5-10; revendications 1-3,5-7 *	1
Y	---	3
Y	IEEE PROCEEDINGS OF THE CONFERENCE OF COMPUTER VISION AND PATTERN RECOGNITION, Washington, 19-23 juin 1983, pages 291-296; HIROSHI MAKINO: "Representation and segmentation of document images" * Page 291, colonne de droite, lignes 30-41; page 293, colonne de gauche, lignes 14-19 * ---	3
A	PATENT ABSTRACTS OF JAPAN, vol. 4, no. 134 (P-28)[616], 19 septembre 1980; & JP-A-55 87 160 (RICOH K.K.) * Abrégé * -----	4-6
		DOMAINES TECHNIQUES RECHERCHES (Int. CL5)
		G 06 F 15/38 G 06 K 9/20
Date d'achèvement de la recherche		Examinateur
09-10-1990		GONZALEZ ORDONEZ O.
CATEGORIE DES DOCUMENTS CITES		
X : particulièrement pertinent à lui seul Y : particulièrement pertinent en combinaison avec un autre document de la même catégorie A : pertinent à l'encontre d'au moins une revendication ou arrière-plan technologique général O : divulgation non-écrite P : document intercalaire		
T : théorie ou principe à la base de l'invention E : document de brevet bénéficiant d'une date antérieure à la date de dépôt et qui n'a été publié qu'à cette date de dépôt ou qu'à une date postérieure. D : cité dans la demande L : cité pour d'autres raisons & : membre de la même famille, document correspondant		